IARS' International Research Journal
ISSN: 2202-2821
ISSN: 1839-6518
iars.researth@gmail.com
International Association of Research Scholars
Organismo Internacional

EARLY DETECTION OF FACTORS, INCLUDING PANDEMICS AND DISASTERS, LEADING TO LANGUAGE ENDANGERMENT: THINKING STATISTICALLY

Peer Reviewed Research Manuscript

EARLY DETECTION OF FACTORS, INCLUDING PANDEMICS AND DISASTERS, LEADING TO LANGUAGE ENDANGERMENT: THINKING STATISTICALLY

Sharma Deepak
Institute of innovation, Technology and Management, India

IARS' International Research Journal, vol. 11, núm. 1, 2021

International Association of Research Scholars, Organismo Internacional

Recepción: 08 Noviembre 2020
Aprobación: 11 Enero 2021
Publicación: 08 Febrero 2021

Redalyc: https://www.redalyc.org/articulo.oa?id=663872611006

Abstract: The target of this research work is to use a statistical technique on different languages to identify significant factors of endangered languages with similar characteristics to build a model for language endangerment. Factor analysis is used to identify factors. The factors are used to construct a model with and without interaction terms. First three variables (i.e. speakers, longitude and latitude) are analyzed to identify two factors and then these three variables and three interaction terms are used to construct the model. Different variables were identified and a model with and without interaction terms is built using the identified factors. The result shows that the model has significant predictive power. The predictors were retrieved from the dataset. The outcome encourages future studies towards defining techniques of language endangerment prediction for analyzing factors of language endangerment.

Keywords: Computational Intelligence, Language Endangerment, Computing, Testing.

I. INTRODUCTION

Statistical modeling for language endangerment is a process, which intends to find the dialect to complete tractability and strength in endangered languages (Trudgill, 2008). This paper presents a comprehensive view of modeling techniques based on computational intelligence and statistical test to find out if language endangerment problem is presented. An endangered language is a dialect that is at danger of dropping out of utilization as its speakers move to talking another language (Axelrod, 2006). It happens when the language has no more local speakers and the language becomes a "dead language". If nobody uses the language it turns into a "wiped out language phase" by (Austin and Sallabank, 2011). Despite the fact that dialects have dependably turned out to be wiped out all through mankind's history, they are as of now vanishing at a very fast pace due to globalization and neocolonialism, and the financially capable dialects command different dialects (Paraschakis, 2013). In this research, we try to elaborate some problems like how to model computation and statistics in language endangerment and their solutions, which are based on artificial intelligence (Hamari, 2014) and (Benjamin, 2014). The paper presents an overview of the usage of computational modeling and artificial intelligence techniques in language endangerment (henceforth LE).

II. Experimental Design and Method

The term Factor Analysis (FA) is a data reduction technique consisting of set of procedures or mechanisms which are used to reduce the many variable data into fewer variables, also termed factors (Johnson and Wichern 2002; Hair et al. 2010; Lawley and Maxwell 1973; Basilevsky, 1981). The method of factorization takes place according to the relevance of characteristics among predictors.

III. Proposed model

The proposed model is designed to identify the influential factors for the language endangerment (LE) problem. The model consists of different phases in order to identify the significant factors for the problem. Figure 1 describes the phases of model testing for LE.

IV. Results and Discussions

The results clearly indicate that the factor variables are far better than the normal predictors. In the case of a superset of interaction terms the results are significant as compared to the other results. The main finding is that in the case of the interaction module values of total variance is good with superset of 6 factor variables as compared to the values of superset of 3 terms. Therefore, we conclude that after the reduction of variables, the values of total variance is also significant. Factor analysis is therefore a good technique for LE problems. In this research, we perform an analysis with review of past studies to check and assess the performance of the Factor analysis statistical technique for LE.

A. Experiment 1

First, a rigorous study is done by following a systematic review, and a few studies (see Austin, 2011; Benjamin, 2014). Secondly, the dataset (Johnson and Wichern, 2002), (Hair et al. 2010), (Lawley, 1973), (Basilevsky, 1981) and (Kaggle, 2016) considered for this study and checked by using the Kaiser Meyer Olkin (KMO) and Bartlett's Test (measures the strength of relationship among the variables) is presented in Table 1 with its measures from Tables 2-6 in experiment 1 with Figure 2 and from Tables 7-12 in experiment number 2 with Figure 3 for the applicability of factor analysis. After investigating this we applied it on the dataset. The significant variables for LE are predicted by factor analysis.

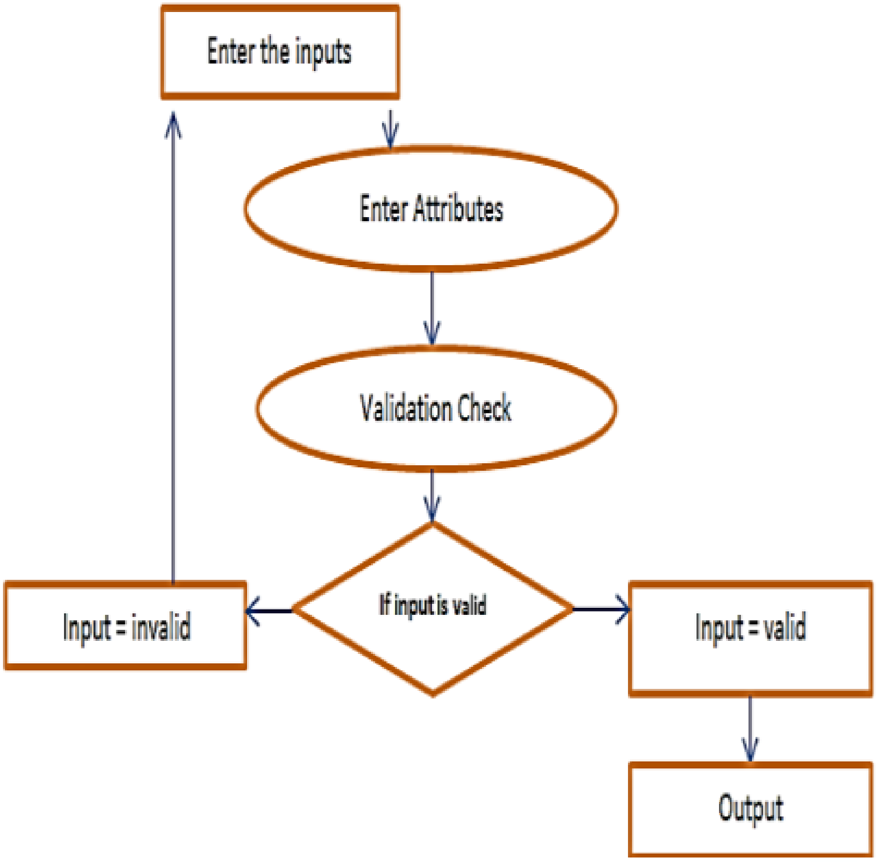


Figure 1:
Describes the phases of model testing for LE

Then, we compare the Total variance (Table 3) which shows the variance explained by each factor and the rotated factor matrix indicates the group of variables. Finally, we summarized the main findings obtained from the results predicted in tabular form.

Table 1

Table 1 KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.484
Bartlett's Test of Sphericity	Approx. Chi-Square	168.635
	Df	3
	Sig.	.000

Table 2

Table 2 Communalities		
	Initial	Extraction
Latitude	1.000	.634
Longitude	1.000	.729
Speakers	1.000	.894

Table 3

Table 3 Total Variance Explained (Extraction Method: Principal Component Analysis.)									
Com.	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.249	41.625	41.625	1.249	41.625	41.625	1.209	40.306	40.306
2	1.008	33.596	75.221	1.008	33.596	75.221	1.047	34.915	75.221
3	.743	24.779	100.000						

The factor analysis has admissible prediction competence for judging LE. The cumulative percentage in total variance explained in experiment 1 is 75.221 and in experiment 2 is 65.564, which are good.

Table 4

Table 4 Component Matrix[a]		
	Component	
	1	2
Latitude	.796	
Longitude	-.678	
Speakers		.859

Extraction Method: Principal Component Analysis. a. 2 components extracted.

Table 5 y 6

Table 5 Rotated Component Matrix[a]

	Component	
	1	2
Latitude	-.721	
Longitude	.830	
Speakers		.945

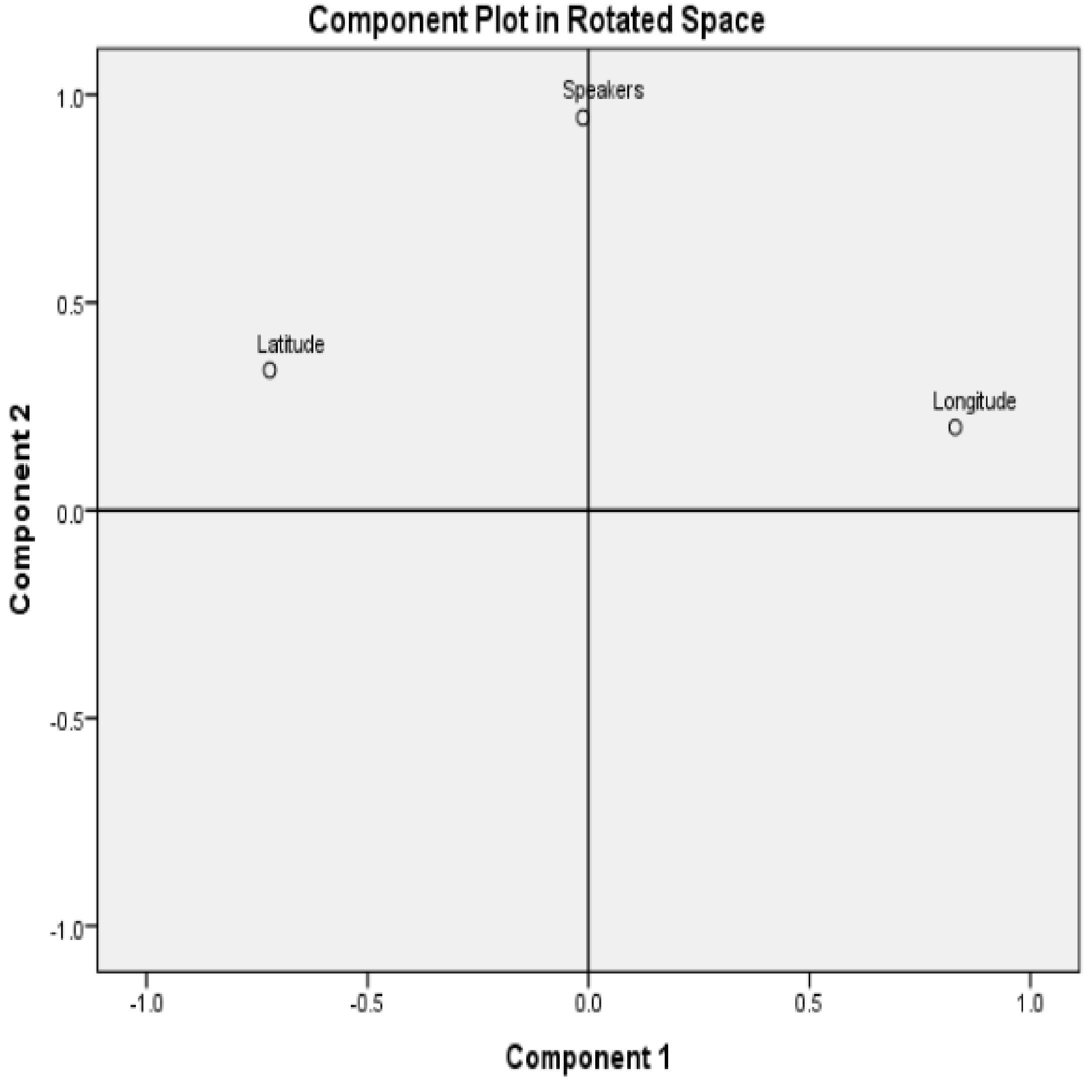
Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 3 iterations.

Table 6 Component Transformation Matrix

Component	1	2
1	-.914	.405
2	.405	.914

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Figure 2:
Scree plot of the eigenvalues of factors i.e. Latitude, Longitude and Speakers



B. Experiment 2

Table 7:
KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.599
Bartlett's Test of Sphericity	Approx. Chi-Square	6745.133
	Df	15
	Sig.	.000

Table 8

Communalities

	Initial	Extraction
Latitude	1.000	.213
Longitude	1.000	.766
Speakers	1.000	.862
A	1.000	.620
B	1.000	.890
C	1.000	.583

Table 9

Table 9 Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.424	40.399	40.399	2.424	40.399	40.399	2.418	40.292	40.292
2	1.510	25.164	65.564	1.510	25.164	65.564	1.516	25.272	65.564
3	.959	15.987	81.551						
4	.569	9.481	91.032						
5	.463	7.711	98.743						
6	.075	1.257	100.000						

Extraction Method: Principal Component Analysis.

Table 10

Table 10 Component Matrix[a]

	Component	
	1	2
Latitude		
Longitude		.867
Speakers	.922	
A		.759
B	.936	
C	.760	

Extraction Method: Principal Component Analysis.

2 components extracted.

Table 11

Table 11 Rotated Component Matrix[a]

	Component	
	1	2
Latitude		
Longitude		.874
Speakers	.928	
A		.774
B	.943	
C	.752	

Table 12

Table 12 Component Transformation Matrix

Component	1	2
1	.996	.084
2	-.084	.996

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

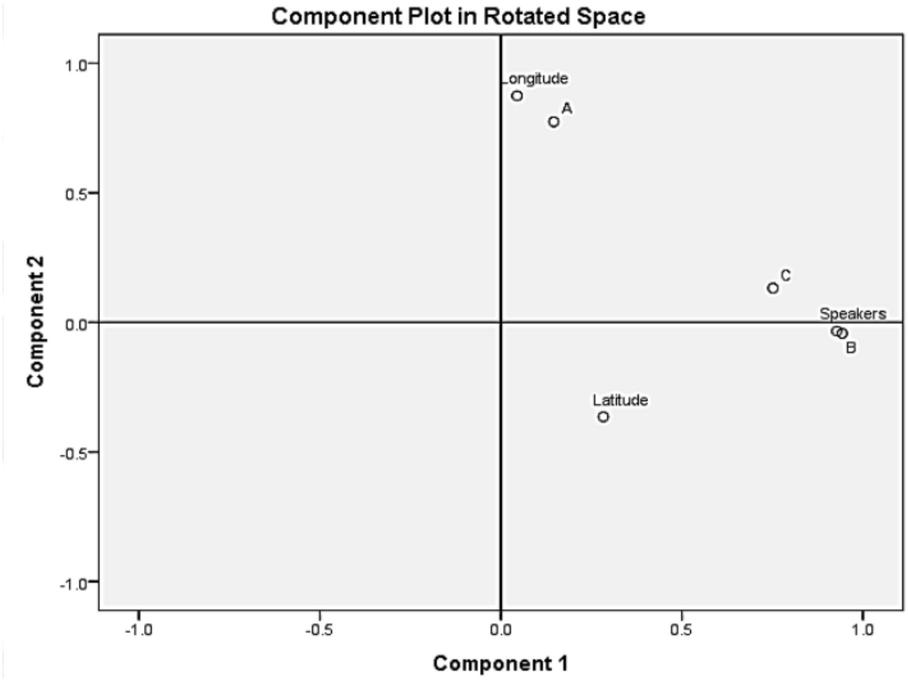


Figure 3:

Scree plot of the eigen values of factors i.e. Latitude, Longitude and Speakers with their interactions

V. Conclusion

This article presents an empirical validation of statistical approaches for LE problem. By using such approaches for LE, we tried to search for the major important factors related to the LE problem. It includes a conceptual discussion of all such methodologies, looking at different criteria of classification and earlier efforts to develop categories for effective and efficient testing for building models of LE. This study has been the basis to develop a proposal for a new anatomy, which is a helpful conceptual tool to both understand and organize the existing work, and to identify possible areas for future research. The study also includes an exhaustive review of the literature in the area, starting from the pioneering works in statistical techniques with testing techniques for LE. The main characteristics of the techniques engaged, as well as the application problems, future directions and results obtained, are presented and can be a source of inspiration for future research in the field.

VI. REFERENCES

Austin, P. K., & Sallabank, J. (Eds.) The Cambridge handbook of endangered languages. Cambridge University Press. (2011).

Axelrod, Robert. Agent-based modeling as a bridge between disciplines. In Leigh Tesfatsion, Kenneth L. Judd (eds.), Handbook of Computational Economics Vol. 2: Agent-Based Computational Economics, 1565–1584. Amsterdam: North Holland/Elsevier. (2006).

Basilevsky, A. Factor analysis regression. Canadian Journal of Statistics, 9(1), 109-117. (1981).

Benjamin, M., & Radetzky, P. Multilingual lexicography with a focus on less-resourced languages: Data mining, expert input, crowdsourcing, and gamification. In 9th edition of the Language Resources and Evaluation Conference (No. CONF). (2014).

"Extinct Languages", Kaggle.com, [Online]. Available: https://www.kaggle.com/the- guardian/extinct-languages. (2010).

Hair, J. F., Anderson, R. E., Babin, B. J., & Black, W. C. Multivariate data analysis: A global perspective. (vol. 7). Pearson Upper Saddle River, New Jersey, Pearson. (2010).

Hamari, J., Koivisto, J., & Sarsa, H. Does gamification work?--a literature review of empirical studies on gamification. In 2014 47th Hawaii international conference on system sciences (pp. 3025-3034). IEEE. 2014

Johnson, R. A., & Wichern, D. W. Applied multivariate statistical analysis (Vol. 5, No. 8). Upper Saddle River, NJ: Prentice hall. (2002).

Lawley, D. N., & Maxwell, A. E. Regression ana factor analysis. Biometrika, 60(2), 331-338. (1973). Online available (https://ieeexplore.ieee.org/abstract/document/675897 8). (2014, January).

Paraschakis, D. Crowd sourcing cultural heritage metadata through social media gaming. Online available (https://muep.mau.se/handle/2043/16114). (2013).

Trudgill, P. On the role of children, and the mechanical view: A rejoinder. Language in Society, 37(2), 277-280. (2008).